1. Let $n=4$, prove $X^TX = 2I$ for the X defined in (3) and the x_i taken at the points specified in (4).

Ans: ∵是 X^TX ∴結果一定是對稱矩陣。

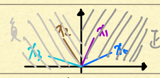
假設結果矩陣是 $A_{4\times4} = X^TX = \begin{pmatrix} a & e & f & g \\ e & b & h & i \\ f & h & c & j \\ g & i & j & d \end{pmatrix}$

因為

$$X = \begin{pmatrix} \frac{1}{\sqrt2} & \cos x_0 & \cos 2x_0 & \cos 3x_0 \\ \frac{1}{\sqrt2} & \cos x_1 & \cos 2x_1 & \cos 3x_1 \\ \frac{1}{\sqrt2} & \cos x_2 & \cos 2x_2 & \cos 3x_2 \\ \frac{1}{\sqrt2} & \cos x_3 & \cos 2x_3 & \cos 3x_3 \end{pmatrix}, \quad X^T = \begin{pmatrix} \frac{1}{\sqrt2} & \frac{1}{\sqrt2} & \frac{1}{\sqrt2} & \frac{1}{\sqrt2} \\ \cos x_0 & \cos x_1 & \cos x_2 & \cos x_3 \\ \cos 2x_0 & \cos 2x_1 & \cos 2x_2 & \cos 2x_3 \\ \cos 3x_0 & \cos 3x_1 & \cos 3x_2 & \cos 3x_3 \end{pmatrix}, \quad \left\{\begin{aligned} x_0 &= \frac{\pi}{8} \\ x_1 &= \frac{3}{8}\pi \\ x_2 &= \frac{5}{8}\pi \\ x_3 &= \frac{7}{8}\pi \end{aligned}\right\} \quad \left\{\begin{aligned} 2x_0 &= \frac{\pi}{4} \\ 2x_1 &= \frac{3}{4}\pi \\ 2x_2 &= \frac{5}{4}\pi \\ 2x_3 &= \frac{7}{4}\pi \end{aligned}\right\}, \quad \left\{\begin{aligned} 3x_0 &= \frac{3}{8}\pi = \frac{3}{8}\pi \\ 3x_1 &= \frac{9}{8}\pi = \pi + \frac{\pi}{8} \\ 3x_2 &= \frac{15}{8}\pi = \pi + \frac{7}{8}\pi \\ 3x_3 &= \frac{21}{8}\pi = 2\pi + \frac{5}{8}\pi \end{aligned}\right\}$$

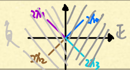
所以 $a = \frac{1}{\sqrt2} \cdot \frac{1}{\sqrt2} \times 4 = 2$

$$e = \frac{1}{\sqrt2}\left(\cos x_0 + \cos x_1 + \cos x_2 + \cos x_3\right) = 0$$



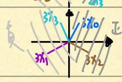
x_0和x_3的cos值正負抵消
x_1和x_2的cos值正負抵消

$$f = \frac{1}{\sqrt2}\left(\cos 2x_0 + \cos 2x_1 + \cos 2x_2 + \cos 2x_3\right) = 0$$



$2x_0$和$2x_1$的cos值正負抵消
$2x_2$和$2x_3$的cos值正負抵消

$$g = \frac{1}{\sqrt2}\left(\cos 3x_0 + \cos 3x_1 + \cos 3x_2 + \cos 3x_3\right) = 0$$



$3x_0$和$3x_3$的cos值正負抵消
$3x_1$和$3x_2$的cos值正負抵消

$$b = \cos^2 x_0 + \cos^2 x_1 + \cos^2 x_2 + \cos^2 x_3$$

$$= \frac{1+\cos 2x_0}{2} + \frac{1+\cos 2x_1}{2} + \frac{1+\cos 2x_2}{2} + \frac{1+\cos 2x_3}{2} = 2 + \frac{1}{2}\underbrace{\left(\cos 2x_0 + \cos 2x_1 + \cos 2x_2 + \cos 2x_3\right)}_{\text{從上面f的推導知,是"0"}} = 2$$

$$h = \cos x_0 \cos 2x_0 + \cos x_1 \cos 2x_1 + \cos x_2 \cos 2x_2 + \cos x_3 \cos 2x_3$$

$$= \frac{\cos(x_0+2x_0)+\cos(x_0-2x_0)}{2} + \frac{\cos(x_1+2x_1)+\cos(x_1-2x_1)}{2} + \frac{\cos(x_2+2x_2)+\cos(x_2-2x_2)}{2} + \frac{\cos(x_3+2x_3)+\cos(x_3-2x_3)}{2}$$

$$= \frac{1}{2}\underbrace{\left(\cos 3x_0 + \cos 3x_1 + \cos 3x_2 + \cos 3x_3\right)}_{\text{by 上面的 g 知,是"0"}} + \frac{1}{2}\left(\cos(-x_0) + \cos(-x_1) + \cos(-x_2) + \cos(-x_3)\right) = 0$$

∵ cos是偶函數 ∴ $\cos(-x) = \cos x$
∴ 這串就是 $\cos x_0 + \cos x_1 + \cos x_2 + \cos x_3$
by 上面的 e, 知是"0"

$$i = \cos x_0 \cos 3x_0 + \cos x_1 \cos 3x_1 + \cos x_2 \cos 3x_2 + \cos x_3 \cos 3x_3$$

$$= \frac{\cos(x_0+3x_0)+\cos(x_0-3x_0)}{2} + \frac{\cos(x_1+3x_1)+\cos(x_1-3x_1)}{2} + \frac{\cos(x_2+3x_2)+\cos(x_2-3x_2)}{2} + \frac{\cos(x_3+3x_3)+\cos(x_3-3x_3)}{2}$$

$$= \frac{1}{2}\left(\overbrace{\cos 4x_0 + \cos 4x_1 + \cos 4x_2 + \cos 4x_3}^{\text{四個都是"0"}}\right) + \frac{1}{2}\underbrace{\left(\cos(-2x_0) + \cos(-2x_1) + \cos(-2x_2) + \cos(-2x_3)\right)}_{\text{by 偶函數 \& 上面的 f, 知是"0"}} = 0$$

$$c = \cos^2(2x_0) + \cos^2(2x_1) + \cos^2(2x_2) + \cos^2(2x_3)$$

$$= \frac{1+\cos 4x_0}{2} + \frac{1+\cos 4x_1}{2} + \frac{1+\cos 4x_2}{2} + \frac{1+\cos 4x_3}{2} = 2 + \frac{1}{2}\underbrace{\left(\cos 4x_0 + \cos 4x_1 + \cos 4x_2 + \cos 4x_3\right)}_{\text{從上面 i 的推導知,是"0"}} = 2$$

$$j = \cos 2x_0 \cos 3x_0 + \cos 2x_1 \cos 3x_1 + \cos 2x_2 \cos 3x_2 + \cos 2x_3 \cos 3x_3$$

$$= \frac{\cos(2x_0+3x_0)+\cos(2x_0-3x_0)}{2} + \frac{\cos(2x_1+3x_1)+\cos(2x_1-3x_1)}{2} + \frac{\cos(2x_2+3x_2)+\cos(2x_2-3x_2)}{2} + \frac{\cos(2x_3+3x_3)+\cos(2x_3-3x_3)}{2}$$

$$= \frac{1}{2}\underbrace{\left(\cos 5x_0 + \cos 5x_1 + \cos 5x_2 + \cos 5x_3\right)}_{\text{和e、f、g相同方法,知是"0"}} + \frac{1}{2}\underbrace{\left(\cos(-x_0) + \cos(-x_1) + \cos(-x_2) + \cos(-x_3)\right)}_{\text{by 偶函數 \& 上面的 i, 知是"0"}} = 0$$

$$d = \cos^2(3x_0) + \cos^2(3x_1) + \cos^2(3x_2) + \cos^2(3x_3)$$

$$= \frac{1+\cos 6x_0}{2} + \frac{1+\cos 6x_1}{2} + \frac{1+\cos 6x_2}{2} + \frac{1+\cos 6x_3}{2} = 2 + \frac{1}{2}\overbrace{\underline{\left(\cos 6x_0 + \cos 6x_1 + \cos 6x_2 + \cos 6x_3\right)}}^{\text{和e、f、g相同方法，知道“0”}} = 2$$

故 $A = X^TX = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = 2I$ 得證 #

2、For $n=4$, Consider the least squares problem $\min\limits_{(a_1,a_2)} \sum\limits_{i=0}^{3} \left(a_1\cos(x_i) + a_2\cos(2x_i) - y_i \right)^2$, where x_i are taken at the points specified in (4). Show the solution is $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \frac{1}{2} X[:,2:3]^T y$, which are the same as the solution of a_1 and a_2 in question 1. Why is it different from the solution of normal equation? (You may reference Theorem 5.5.6)

Ans: $\min\limits_{(a_1,a_2)} \sum\limits_{i=0}^{3} \left(a_1\cos(x_i) + a_2\cos(2x_i) - y_i \right)^2$，其實就是解 $y = a_1\cos x + a_2\cos 2x$ 在 $\left\{ \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} \right\}$ 的 least squares problem.

$$\Rightarrow \begin{pmatrix} \cos x_0 & \cos 2x_0 \\ \cos x_1 & \cos 2x_1 \\ \cos x_2 & \cos 2x_2 \\ \cos x_3 & \cos 2x_3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad \Rightarrow \underbrace{\begin{pmatrix} \cos x_0 & \cos x_1 & \cos x_2 & \cos x_3 \\ \cos 2x_0 & \cos 2x_1 & \cos 2x_2 & \cos 2x_3 \end{pmatrix} \begin{pmatrix} \cos x_0 & \cos 2x_0 \\ \cos x_1 & \cos 2x_1 \\ \cos x_2 & \cos 2x_2 \\ \cos x_3 & \cos 2x_3 \end{pmatrix}}_{\overset{\text{第一小題的 } b,h,h,c}{=} \begin{pmatrix} b & h \\ h & c \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \cos x_0 & \cos x_1 & \cos x_2 & \cos x_3 \\ \cos 2x_0 & \cos 2x_1 & \cos 2x_2 & \cos 2x_3 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\Rightarrow 2\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \cos x_0 & \cos x_1 & \cos x_2 & \cos x_3 \\ \cos 2x_0 & \cos 2x_1 & \cos 2x_2 & \cos 2x_3 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{1}{2} \underbrace{\begin{pmatrix} \cos x_0 & \cos x_1 & \cos x_2 & \cos x_3 \\ \cos 2x_0 & \cos 2x_1 & \cos 2x_2 & \cos 2x_3 \end{pmatrix}}_{= X[:,2:3]^T} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{1}{2} X[:,2:3]^T y$$，第一小題得證 #

至於第二小題，題目說然是問為何不同，但實際上兩者的 a_1、a_2 就是相同，只不過 $\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$ 兩者不同，因為本小題的 a_0 和 a_3 均為0。原因如下敘述：

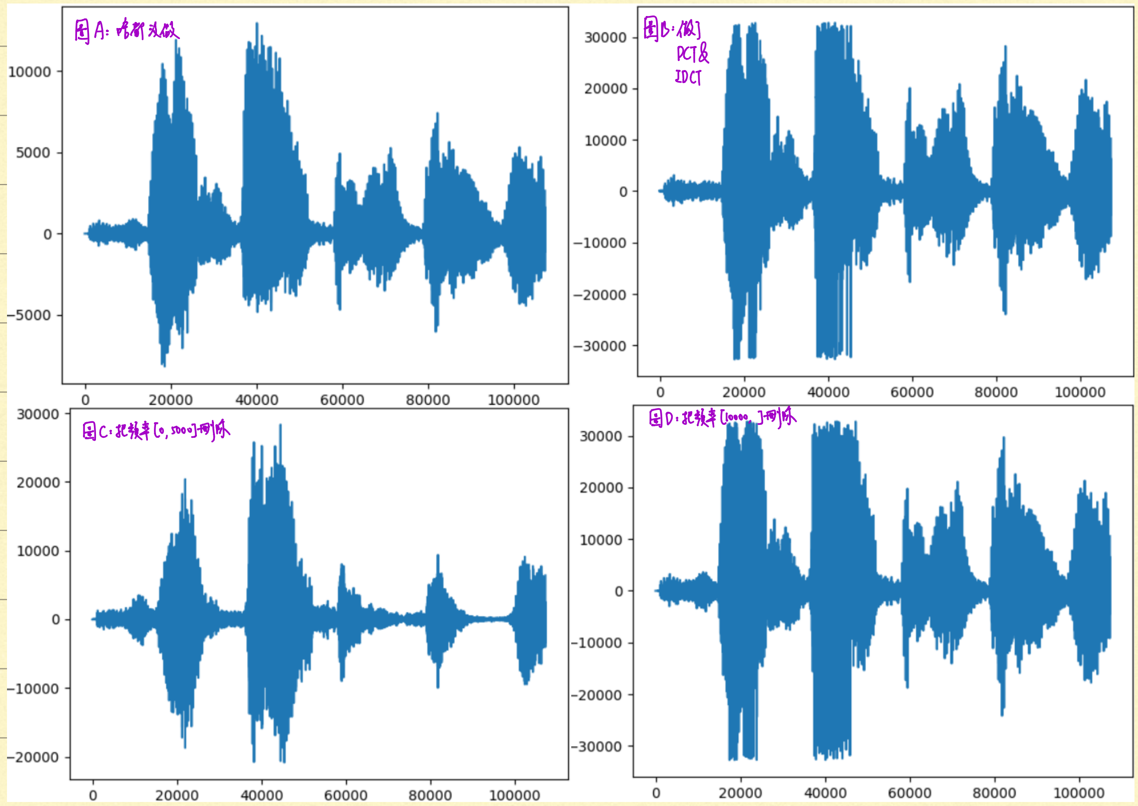
由第一題我們得知 $X^TX = 2I$，所以若是用第一題的 normal equation $X^TXa = X^Ty \Rightarrow 2Ia = X^Ty \Rightarrow a = \frac{1}{2}X^Ty$。

若將 $Xa=y$ 的 a_0 和 a_3 設為0，則 $\begin{pmatrix} \frac{1}{\sqrt{2}} & \cos x_0 & \cos 2x_0 & \cos 3x_0 \\ \frac{1}{\sqrt{2}} & \cos x_1 & \cos 2x_1 & \cos 3x_1 \\ \frac{1}{\sqrt{2}} & \cos x_2 & \cos 2x_2 & \cos 3x_2 \\ \frac{1}{\sqrt{2}} & \cos x_3 & \cos 2x_3 & \cos 3x_3 \end{pmatrix} \begin{pmatrix} 0 \\ a_1 \\ a_2 \\ 0 \end{pmatrix}$，完全就是第二題中的第一小題的 least squares problem 所用的 normal equation。

又因為前面推得的 $a=\frac{1}{2}X^Ty$，得知 $\begin{pmatrix} 0 \\ a_1 \\ a_2 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{2}} & \cos x_0 & \cos 2x_0 & \cos 3x_0 \\ \frac{1}{\sqrt{2}} & \cos x_1 & \cos 2x_1 & \cos 3x_1 \\ \frac{1}{\sqrt{2}} & \cos x_2 & \cos 2x_2 & \cos 3x_2 \\ \frac{1}{\sqrt{2}} & \cos x_3 & \cos 2x_3 & \cos 3x_3 \end{pmatrix}^T \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix}$

$$故 \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{1}{2} X[:, 2:3]^T y, \text{ 小題證畢}$$

3. Plot the waves before and after the processing, and compare their differences.



比較圖A和圖B，發現：如果對於一個合成聲波用傅立葉轉換(Fourier Transform)，把time domain轉為frequency domain，之後再用傅立葉反轉換(Inverse Fourier Transform)，把frequency domain再轉為time domain，其實啥都不做v.s.轉換過去再轉換回來，圖形是一樣的，但振幅略有不同。

比較圖A和圖C，發現：把頻率[0:5000]給過濾掉之後，圖C中第三四五個脈波的圖形幾乎被削減到快沒了，第一個脈波的開頭處也大幅度被削弱，所以暫可推測圖A的聲音原檔裡，頻率[0:5000]的比例應該非常高

比較圖A和圖D，發現：把頻率[10000:]給過濾掉之後，圖D中每個脈波的振幅都比原本圖A來得高。因為圖A作為原圖，是把所有頻率和振幅綜合考慮計算，所以綜合兩圖來看，可以推測，因為把[10000:]的頻率都刪掉了，所以原本統計分析的不合群的那些造成相關係數變小的頻率都被刪掉了，所以圖D才感覺振幅都變大。

綜合
(1)圖A和圖C的比較，以及
(2)圖A和圖D的比較，可以結論：
原本的聲音波形，大多是由頻率[0:5000]組成。

4. Discuss the effects on the sound signals after you remove the low frequency part, and the effects on the sound signals after you remove the high frequency part.

由第三題的圖C和和圖D所生成的聲音檔案，可以聽到：
若過濾掉低頻部分([0:5000])，則音高較高且聽起來尖銳，且有雜音
若過濾掉高頻部分([10000:0])，則音高較低，且有雜音

```
# do some filtering
#data0_1[0:5000] = 0
#data0_1[10000:] = 0
```

過濾頻率的code片段

5. Explain your algorithm

```
## Added code by myself
average0_1 = np.mean(np.abs(data0_1))
average1_1 = np.mean(np.abs(data1_1))
average2_1 = np.mean(np.abs(data2_1))
average3_1 = np.mean(np.abs(data3_1))


# do some filtering
threshold0_1 = average0_1 * 0.95;
temp0_1 = []; countZeroes0_1 = 0
for i in range(len(data0_1)):
    if abs(data0_1[i]) <= threshold0_1:
        countZeroes0_1 += 1
    else:
        if countZeroes0_1 != 0:
            temp0_1.append(0)
            temp0_1.append(countZeroes0_1)
            countZeroes0_1 = 0
        temp0_1.append(data0_1[i])
if countZeroes0_1 != 0:
    temp0_1.append(0)
    temp0_1.append(countZeroes0_1)
done0_1 = np.array(temp0_1)
##############################
```

壓縮的部分：
因為不同的聲音檔案振幅的範圍可大可小，如果只用一個常數當作篩選門檻可能過於武斷且缺乏彈性，所以我想根據得到的聲音檔案計算所有頻率的振幅的平均。但因為有正負振幅的差別，為了取平均的時候不要算成接近0的數字導致啥都沒被篩掉，所以先把所有振幅取絕對值。

之後我根據這個平均當作門檻的基準。平均的概念差不多就是PR50左右的意思，但是為了想要留住超過50%的資料，所以我把門檻調低了一點(0.95倍)。之後就是開始根據每個頻率做篩選。
篩選的準則就是：如果一個頻率的振幅小於等於門檻，那我就直接刪掉(a.k.a. 振幅變成0)。如果高過門檻，就保留振幅的精準訊息。
如果振幅高過門檻，因為我需要保留最精確的振幅資訊，所以這部分能對壓縮做操作的空間很小。我們能夠提高壓縮率的，主要就是把一連串的0，記錄起來到底有幾個0。(例如: list=[0, 0, 0, 0, 0]，我就紀錄為五個"0"。紀錄[0, 0, 0, 0, 0]佔用了五格，但是紀錄[5, 0]只占用兩格。)

```
# please implement your own restore function, no global variables are allowed to be use
def restore(compressed):

    temp = []
    idx_for_compressed = 0

    while idx_for_compressed < len(compressed):
        if compressed[idx_for_compressed] != 0:
            temp.append(compressed[idx_for_compressed])
            idx_for_compressed += 1
            continue
        else:
            idx_for_compressed += 1
            num = int(compressed[idx_for_compressed])
            list_of_zeroes_to_append = [0 for i in range(num)]
            temp[len(temp):] = list_of_zeroes_to_append[:]
            idx_for_compressed += 1
            continue
```

解壓縮的部分：
就是根據上面的壓縮的方法，inverse地做。